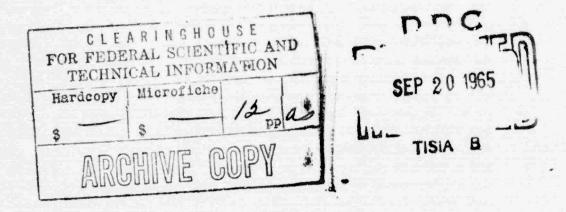# BAYESIAN INFERENCE CONCERNING MANY PARAMETERS,
WITH REFERENCE TO SUPERSATURATED DESIGNS*

F. J. ANSCOMBE
Princeton University†

INTRODUCTION

Real scientific judgments ar not exactly reflected in the theory of consistent behaviour of Ramsey, de Finetti, and Savage, since in the theory the person is postulated to have infinitely fine perceptiveness. When the theory is applied to well-defined statistical problems, the initial probabilities of parameters and the utility function postulated in the calculations are derived from rather imprecise perceptions or judgments. Often, the final results are insensitive to moderate changes in the initial probabilities and the utility function, and it is natural to choose mathematical expressions for these functions that roughly represent the intuitive judgments and are at the same time mathematically convenient to handle. Mistakes are rather easily made, however. That is, one may postulate mathematical forms for the initial probability distribution or for the utility function, under the impression that intuition is thereby represented fairly, when in fact the postulated forms have implications at variance with intuition.

It is quite common in statistics to work with very simple loss functions, notably with quadratic loss functions and also with two-valued loss functions. In many cases these are good enough. But a quadratic loss function is not so

good when associated with a probability distribution having an infinite variance. The trouble here is obvious, and usually there will be no hesitation over taking the obvious remedy, of changing the loss function. It is possible also for there to be a conflict between the data and the initial distribution of parameters. This may not be altogether obvious, though if it is noticed either the data or the initial distribution will have to be rejected or modified. Suppose, as an extreme example, that a person proposes to take readings of the weight μ of an object, and let it be given that the readings will differ from μ by a chance "error" having a normal distribution with zero mean and 1 g. standard deviation. Before the readings are taken, the person may judge that μ will lie between 200 g. and 300 g., and having no particular opinion as to where μ will come in this range he may be incautious enough to postulate a uniform distribution for the initial probability over just this interval, with zero probability outside it. Suppose now that he takes some readings, and finds them to be somewhere near 500 g. If after careful checking he accepts the readings as genuine and trustworthy, and is prepared to conclude that, contrary to prior expectation, μ is indeed in the neighbourhood of 500 g., then his asserted initial distribution must have been wrong. He did not really mean to imply absolute conviction that μ lay between 200 g. and 300 g. If he had had such a conviction, he would have been forced to reject the observations as spurious, or else, by a long stretch of the imagination amounting to lunacy, to have concluded from Bayes's theorem that μ was probably very close to 300 g.

It is customary to express the total probability distribution for an experiment, extending over the Cartesian product of a "sample space" and a "parameter space," by two component distributions, (i) the conditional distribution over the sample space, given any point in the parameter space (this is often called the statistical specification, or class of admissible hypotheses, or "model" for the experiment), and (ii) the marginal distribution over the parameter space, usually referred to simply as the initial or "prior" distribution. Usually, the former is supposed to be a chance distribution with frequency interpretation, whereas the latter represents subjective opinion. It is well known that the observations are capable of contradicting the first type of distribution, as for example when the distribution is asserted to be normal but the observations are strikingly discrepant with the hypothesis of normality. It does not seem to be generally realized that the observations can just as clearly contradict the second type of distribution.

This is liable to happen, in particular, if there are many parameters. Suppose that according to the statistical specification the observations y_i $(i = 1, 2, \ldots, n)$ have normal independent chance distributions with known common variance and mean values expressed as linear functions of some parameters. It is tempting to postulate as the initial distribution for the parameters some particular multivariate normal distribution, because such a distribution is "conjugate" to the statistical specification (Raiffa and Schlaifer [6]). However, by a (fully known) non-singular linear transformation, the parameters can be made to have independent normal initial distributions with zero means and unit variances; and

now—if the number of parameters is large—by the law of large numbers we deduce that a certain linear function of the observations has high probability of being close to zero, and that a certain quadratic function of the observations has high probability of being close to one, and similar statements about higher moments. But these particular functions of the observations may in fact yield values that are far from the predicted values. Perceiving this difficulty, Duncan [3], and possibly also Dunnett [4], have hinted that the initial distribution for the parameters should not be completely postulated in advance, but should be permitted to have one or more adjustable constants in it, which would be chosen, after a preliminary analysis of the observations, to avoid one or more specific conflicts of the above sort. Such an adjusted distribution cannot be said to represent prior opinion directly, since it depends on the observations. Its use may possibly be justifiable, but we may reasonably ask to see the justification.

The main purpose of this paper is to illustrate a more satisfying way of assigning the initial distribution for the parameters, in a problem where there are many parameters. We consider an experiment in which a large number of factors are tested or "screened." It is expected that most of the factors will have small effects, and importance attaches to identifying any factors that have substantial effects. We introduce an initial distribution expressing indifference of opinion concerning the factor responses, but a certain expectation as to the relative magnitudes of the responses in aggregate; such a distribution should appear reasonable (we believe) for some kinds of exploratory experiments. Conditions on the design that facilitate the ensuing analysis are formulated. Particular attention is paid to supersaturated designs, for which the number of parameters exceeds the number of observations.

Our formulation of this problem is based on unpublished work by Beale and Mallows [1], to whom is due the idea of modifying an ordinary least-squares analysis in the light of a suitably chosen prior distribution. On the subject of screening experiments in general, including supersaturated experiments, see Satterthwaite [7] and the associated discussion. Booth and Cox [2] have recently considered the construction of supersaturated designs, under a condition (not imposed here) that each factor may be tested at only two levels. The considerations of this paper have some relation to Stein's work on many-parameter estimation, as for example [8] and [9].

Tiao and Zellner [10] have treated essentially the same problem, in a paper that came to hand after this was written.

A FACTOR-SCREENING EXPERIMENT

Specification and notation

Let $y_1, y_2, \ldots, y_n$ be readings obtained in a factorial experiment in which f factors are varied. Let the following specification be postulated, implying that the effects of the factors are linear and additive:

$$(1) \qquad y_i = \mu + \sum_{r=1}^{f} x_{ir}\beta_r + e_i \qquad (i = 1, 2, \ldots, n).$$

Here the coefficients x_{ir}, representing the levels of the factors, are known and freely adjustable in the planning of the experiment; the mean μ and the response coefficients β_r are parameters whose values are not certainly known; and the errors e_i are realizations of chance variables independently and identically distributed in a normal distribution with mean 0 and variance σ^2.

Let X denote the $n \times (f + 1)$ matrix of coefficients of the parameters μ, β_r. in (1), that is,

$$(2) \qquad X = \begin{pmatrix} 1 & x_{11} & x_{12} & \ldots & x_{1f} \\ 1 & x_{21} & x_{22} & \ldots & x_{2f} \\ . & . & . & & . \\ 1 & x_{n1} & x_{n2} & \ldots & x_{nf} \end{pmatrix}.$$

By redefining μ and the zeros of the factor levels if necessary, we can make every other column of X orthogonal to the first column. Thus we shall suppose

$$(3) \qquad \sum_i x_{ir} = 0 \qquad\qquad (r = 1, 2, \ldots, f).$$

We shall also suppose that X has the greatest possible rank for its size, namely $\min(n, f + 1)$. We shall be interested in large values for n and f, but the formal requirement in what follows is merely that $n > 3$, $f > 1$. If $f \leqslant n - 2$, we say that the design is *unsaturated*; the parameters μ, β_r, can be estimated by the method of least squares, and σ^2 can be estimated from the residual sum of squares. If $f = n - 1$, the design is *saturated*, and the method of least squares yields estimates of μ and β_r, but not of σ^2. If $f > n$, the design is *supersaturated*, and the method of least squares yields a unique estimate for μ but not a unique estimate for β_r.

It will be convenient to consider a reduced form for all but the first column of X. Let T denote an $n \times n$ orthogonal matrix, with entries t_{ij}, such that every entry in the first row is equal to $1/\sqrt{n}$. Then T transforms the observations (y_i) to (u_i), where

$$(4) \qquad u_i = \sum_j t_{ij} y_j,$$

and

$$(5) \qquad u_1 = \sum_j y_j / \sqrt{n} = \bar{y}\sqrt{n}.$$

Because of (3), X is transformed to

$$(6) \qquad TX = \begin{pmatrix} \sqrt{n} & 0 & 0 & \ldots & 0 \\ 0 & w_{21} & w_{22} & \ldots & w_{2f} \\ . & . & . & \ldots & . \\ 0 & w_{n1} & w_{n2} & \ldots & w_{nf} \end{pmatrix}.$$

Let W stand for the $(n - 1) \times f$ matrix of coefficients w_{ir}, $(i = 2, 3, \ldots, n;$ $r = 1, 2, \ldots, f)$. For some purposes, W is a more convenient description of the design than X.

Let C stand for the $f \times f$ matrix $W'W$ (where the prime denotes the transpose), having entries

$$(7) \qquad c_{rs} = \sum_{(i)} w_{ir} w_{is} = \sum_i x_{ir} x_{is}.$$

In summations, the suffixes r, s, will be understood to run always from 1 to f, and the suffixes i, j, to run from 1 to n, unless the suffix is enclosed in parentheses, as above in (7), in which case it will be understood to run from 2 to n.

Purpose

The purpose of conducting the experiment and observing the y_i is to make inferences concerning the response coefficients β_r. We shall express the results of the experiment primarily in the form of a posterior distribution for the β_r. We shall also consider the particular problem of selecting from among the factors any for which the corresponding response coefficient is substantially different from zero. This problem will be characterized by the following loss function.

We suppose that, associated with the rth factor, there is a given number h_r, serving as a threshold of importance, such that if $|\beta_r|$ were known to exceed h_r we should prefer to classify the rth factor as "interesting" and retain it for future investigation, whereas if $|\beta_r|$ were known to be less than h_r we should prefer to give the verdict "uninteresting" and discard the factor. Let the utility loss in discarding the factor be supposed equal to β_r^2, and the cost of retaining the factor for further investigation be supposed independent of β_r, and therefore equal to h_r^2. Then if the value of β_r is not known certainly, the decision to retain the factor will be preferred if

$$(8) \qquad E(\beta_r^2) > h_r^2,$$

and the decision to discard will be preferred if the inequality is reversed. Here the expectation is with respect to the available marginal probability distribution for β_r.

Thus in addition to determining a joint posterior distribution for all the response coefficients β_r, given the results of the experiment, we shall be interested in the value of $E(\beta_r^2)$ for each r, with respect to this posterior distribution.

Initial probability distribution for the parameters

We have to supply a joint initial distribution for the parameters, μ, σ^2, and the β_r. For μ we shall postulate the uniform distribution over the whole real line, independent of the distribution for the other parameters. Our procedure of analysis will therefore be invariant under changes in the origin of measurement of the y_i. For σ^2 it will be simplest to consider the possibility that σ^2 is certainly known. But if σ^2 is not known certainly it will be supposed to have the following initial-probability element, independent of the distribution of the other parameters:

$$(9) \qquad (\sigma^{-2})^{k/2-1} \exp[-a^2/2\sigma^2] d(\sigma^{-2}),$$

where k and a are given positive numbers. (There is no need to insert a normalizing factor, since it will cancel out when we use Bayes's theorem.)

As for the response coefficients β_r, we wish to express a prior belief that most of them will be small. If f is large, no joint normal distribution for the β_r will do, because of the law-of-large-numbers effect, as already mentioned.

Observe first that if all the entries in any column of W (or in any column of X other than the first) are multiplied by a non-zero constant and the corresponding β divided by that constant, the problem is essentially unchanged. Let us suppose the columns have been so adjusted that the initial opinion concerning the magnitude of each β is the same. The set of β's, if we could observe them, would look like a sample from some population. Let us consider the possibility of postulating that the population would be normal and have zero mean.

The suggestion that initially each β has zero expectation and that the joint distribution of the β's is invariant under permutations can be given a certain sort of objective validity by performing two acts of randomization: (a) each column of W (or each column of X other than the first), together with the corresponding β, is either unchanged or multiplied by -1, according to the flip of a fair coin, choices being independent for different columns; (b) the f factors are given a random permutation before being numbered from 1 to f.

These randomizations do no harm, and, like other more usual randomizations, are probably advisable as a protection against unconscious biases and specification errors if a series of similar experiments is to be performed. But after the results of the randomizations are known, the fact of randomization has no bearing on the plausibility of a normal population of β's with zero mean. The normality of the distribution is open to empirical refutation, after a large number of β's have been estimated, just as the assumption that the e's have a normal distribution is open to empirical refutation, provided there is enough replication in the experiment. Both normality assumptions are logically alike and may be regarded as belonging to the "specification" rather than to the "initial distribution" part of the total assumed probability distribution for the experiment. We then have only three parameters for which we have to assert initial distributions, namely μ and σ^2, as already discussed, and also τ^2, the variance of the normal population from which the β's are supposed drawn. For the latter let us postulate the initial-probability element

$$(10) \qquad (\tau^{-2})^{l/2-1} \exp[-b^2/2\tau^2] d(\tau^{-2}),$$

where l and b are given positive numbers. The mathematical forms for (9) and (10) have, of course, been chosen for mathematical convenience. Each distribution may be adjusted to have any mean and any variance.

Thus the total probability element for the experiment is proportional to

$$(11) \quad \exp\left[- \sum_i \left(y_i - \mu - \sum_r x_{ir}\beta_r\right)^2 \Big/ 2\sigma^2 - \sum_r \beta_r^2/2\tau^2 - b^2/2\tau^2\right]$$
$$(\tau^{-2})^{(l+n)/2-1} \prod_i dy_i \prod_r d\beta_r d\mu d(\tau^{-2}),$$

if σ^2 is known certainly. If σ^2 is not known certainly but has the initial distribution (9), then the total probability element is proportional to (11) multiplied by the further factor

(12)
$$(\sigma^{-2})^{(1+n)/2-1}\exp[-a^2/2\sigma^2]d(\sigma^{-2}).$$

But for the present let us suppose σ^2 known.

If we integrate (11) with respect to τ^2, we obtain the marginal distribution for all the other variables. The probability element is proportional to the product of the following two factors, which we separate for ease of discussion:

(13)
$$\exp\left[-\sum_i \left(y_i - \mu - \sum_r x_{ir}\beta_r\right)^2 \Big/ 2\sigma^2\right] \prod_i dy_i d\mu,$$

and

(14)
$$\left(b^2 + \sum_r \beta_r^2\right)^{-(l+f)/2} \prod_r d\beta_r.$$

Thus our assumptions about the β_r, that they are a sample from a normal population having zero mean and variance τ^2, where τ^2 has the initial distribution (10), is equivalent to postulating (14) as the joint initial distribution for the β_r. The distribution (14) has the following properties. The marginal distribution for any one response coefficient, say β_1, is proportional to

(15)
$$(b^2 + \beta_1^2)^{-(l+1)/2}d\beta_1,$$

and this has a great dispersion if l is small. But the conditional distribution for β_1, given all the other β's, and given, in particular, that $\beta_2^2 + \beta_3^2 + \ldots + \beta_f^2 = \rho_1^2$, say, is proportional to

(16)
$$(b^2 + \rho_1^2 + \beta_1^2)^{-(l+f)/2}d\beta_1,$$

and this is close to a normal distribution with zero mean and variance

$$(b^2 + \rho_1^2)/(l + f - 1)$$

if f is large. These properties are intuitively satisfactory, and do not resemble those of any joint normal distribution.

INFERENCES FROM THE EXPERIMENT

Generalities

The posterior distribution for the β_r is derived (according to Bayes's theorem) from the total probability distribution for the experiment, (13) and (14), by integrating with respect to the unwanted parameter μ and conditioning on the observed values of the y_i. The factor (14) is not affected by these operations, but (13) becomes

(17)
$$\exp\left[-\sum_i \left(y_i - \bar{y} - \sum_r x_{ir}\beta_r\right)^2 \Big/ 2\sigma^2\right],$$

where y_i and $\bar{y}$ now stand for the observed values. Thus the desired posterior distribution is proportional to the product of (14) and (17).

This is for σ^2 a known constant. If σ^2 is a random variable, the factor (12) must also be reckoned with. The integration with respect to μ changes the n in (12) to $n - 1$, and then it is necessary to integrate also with respect to σ^2.

The resulting desired posterior distribution for the β_r is proportional to the product of (14) and

$$(18) \qquad \left[a^2 + \sum_i \left(y_i - \bar{y} - \sum_r x_{ir}\beta_r\right)^2\right]^{-(k+n-1)/2}.$$

Note that, in regard to the sum of squares appearing in (17) and (18), we have:

$$(19) \qquad \sum_i \left(y_i - \bar{y} - \sum_r x_{ir}\beta_r\right)^2 = \sum_{(i)} \left(u_i - \sum_r w_{ir}\beta_r\right)^2.$$

Let (β_r^*) denote a set of values for (β_r) that minimizes this sum of squares. If the design is saturated or supersaturated, the minimized sum vanishes. (β_r^*) is uniquely determined by the minimization if the design is unsaturated or saturated, i.e. if $f \leqslant n - 1$. For a supersaturated design, we shall impose below a further condition which will determine (β_r^*) uniquely. In any case, the sum of squares (19) can be expressed, from (7), as

$$(20) \qquad \sum_{(i)} \left(u_i - \sum_r w_{ir}\beta_r^*\right)^2 + \sum_{rs} c_{rs}(\beta_r - \beta_r^*)(\beta_s - \beta_s^*).$$

The first half of this expression, the minimized sum of squares, does not involve the β_r, and so, ignoring it, we may replace (17), in the posterior distribution for the β_r when σ^2 is known, by

$$(21) \qquad \exp\left[- \sum_{rs} c_{rs}(\beta_r - \beta_r^*)(\beta_s - \beta_s^*)/2\sigma^2\right].$$

The corresponding expression for (18), in the posterior distribution for the β_r when σ^2 is not known, is

$$(22) \qquad \left[A^2 + \sum_{rs} c_{rs}(\beta_r - \beta_r^*)(\beta_s - \beta_s^*)\right]^{-(k+n-1)/2},$$

where $A^2 = a^2 + \sum_{(i)} (u_i - \sum_r w_{ir}\beta_r^*)^2$. For the rest of this paper, we shall consider σ^2 known and work with (21), but our results can immediately be translated into the corresponding expressions derived from (22).

Under our assumption that X is of maximum rank, C is of rank $\min(n - 1, f)$. Because the prior distribution (14) involves the simple equal-weighted sum $\sum_r \beta_r^2$, the posterior distribution will be easiest to make computations with if the experiment has been so designed that all the non-zero roots of C are equal. The common numerical value of these roots is arbitrary, because all the entries in X, except those in the first column, or equivalently all the entries in W, could be multiplied by a common non-zero constant, and all the β_r and b divided by that constant, and the problem would be unchanged. For definiteness, we shall choose the common value of the roots to be n. This permits the design conditions below to be expressed as succinctly as possible, but with appropriate slight changes in the formulas any other scaling could be used.

Let g be the positive number defined by

(23)
$$g^2 = \sum_r \beta_r^{*2}.$$

We shall assume that β_r^* does not vanish for every r, and hence that g exists.

The case $f \leqslant n - 1$ (unsaturated or saturated design)

The rank of C is f. If all the roots of C are equal to n, C must be n times the identity matrix. We then have

(24)
$$\beta_r^* = \sum_i y_i x_{ir}/n,$$

and ng^2 is equal to the sum of squares for factor responses, of f degrees of freedom, in the ordinary analysis of variance of the observations. The implied condition on X is expressed in the following

DESIGN CONDITION $(f \leqslant n - 1)$. *The columns of X are orthogonal, and the sum of squares of entries in every column is equal to n.*

By an orthogonal transformation we can transform (β_r) to (γ_r), where

(25)
$$\gamma_1 = \sum_r \beta_r^* \beta_r / g$$

and the posterior distribution for the γ_r is proportional to

(26)
$$\frac{\exp[-n\{(\gamma_1 - g)^2 + \gamma_2^2 + \ldots + \gamma_f^2\}/2\sigma^2]\prod_r d\gamma_r}{(b^2 + \gamma_1^2 + \gamma_2^2 + \ldots + \gamma_f^2)^{(1+f)/2}}.$$

This distribution obviously has circular symmetry about the γ_1-axis. Let

(27)
$$\rho_1^2 = \gamma_2^2 + \gamma_3^2 + \ldots + \gamma_f^2.$$

Then the joint posterior distribution for γ_1 and ρ_1 is proportional to

(28)
$$\frac{\exp[-n\{(\gamma_1 - g)^2 + \rho_1^2\}/2\sigma^2]\rho_1^{f-2}d\gamma_1 d\rho_1}{(b^2 + \gamma_1^2 + \rho_1^2)^{(1+f)/2}}.$$

(26) and (28) is our primary answer to the inference problem.

For the particular inference problem of selecting "interesting" factors, we need to evaluate $E(\beta_r^2)$. Inverting the transformation, we can express β_r as a linear form in the γ_s; the sum of squares of the coefficients is equal to 1, and the coefficients of γ_1 is β_r^*/g. We have clearly

(29) $E(\gamma_r \gamma_s) = 0$ for $r \neq s$,

$$E(\gamma_2^2) = E(\gamma_3^2) = \ldots = E(\gamma_f^2) = E(\rho_1^2)/(f - 1).$$

Hence

(30) $E(\beta_r^2) = (\beta_r^*/g)^2 E(\gamma_1^2) + [1 - (\beta_r^*/g)^2]E(\rho_1^2)/(f - 1).$

This can be calculated (for every value of r) as soon as $E(\gamma_1^2)$ and $E(\rho_1^2)$ have been evaluated, and the latter may be found by numerical double integration over the region $-\infty < \gamma_1 < \infty$, $0 < \rho_1 < \infty$, of three expressions in turn, namely (28) as it stands, (28) multiplied by γ_1^2, and (28) multiplied by ρ_1^2. The

first of these is required in order to determine the normalizing factor for the distribution (28).

The case $f > n$ (supersaturated design)

The rank of C is $n - 1$. The non-zero roots of C $(= W'W)$ are equal to those of WW', and if there are $n - 1$ such roots equal to n, WW' must be n times the identity matrix. Thus the rows of W are orthogonal and the sum of squares of entries in every row is equal to n; the same is therefore true for X. A set of characteristic vectors for C, corresponding to the non-zero roots, is the vectors whose components are the rows of W. Let (β_r^*) be the unique solution of the equations

$$(31) \qquad u_i = \sum_r w_{ir}\beta_r^* \qquad (i = 2, 3, \ldots, n)$$

that lies in the linear subspace spanned by these characteristic vectors; we refer to this space as V. The general solution of the equations is then equal to (β_r^*) plus any arbitrary vector orthogonal to V, and so (β_r^*) may be characterized as the solution of (31) that minimizes g. We see easily that β_r^* is given by (24), and

$$(32) \qquad ng^2 = \sum_i (y_i - \bar{y})^2.$$

There is an orthogonal $f \times f$ matrix $P = (p_{rs})$, transforming (β_r) to (γ_r), thus $\gamma_r = \sum_s p_{rs}\beta_s$, such that the first row (p_{1s}) is proportional to (β_s^*), i.e., we have (25), and such that the first $n - 1$ rows of P are linearly dependent on the rows of W, i.e., they span V. Then the posterior distribution for the γ_r is proportional to

$$(33) \qquad \frac{\exp[-n\{(\gamma_1 - g)^2 + \gamma_2^2 + \ldots + \gamma_{n-1}^2\}/2\sigma^2]\prod_r d\gamma_r}{(b^2 + \gamma_1^2 + \gamma_2^2 + \ldots + \gamma_f^2)^{(1+f)/2}}.$$

Let

$$(34) \qquad \rho_1^2 = \gamma_2^2 + \gamma_3^2 + \ldots + \gamma_{n-1}^2, \quad \rho_2^2 = \gamma_n^2 + \gamma_{n+1}^2 + \ldots + \gamma_f^2.$$

Then the joint posterior distribution for γ_1, ρ_1, and ρ_2 is proportional to

$$(35) \qquad \frac{\exp[-n\{\gamma_1 - g)^2 + \rho_1^2\}/2\sigma^2]\rho_1^{n-3}\rho_2^{f-n}d\gamma_1 d\rho_1 d\rho_2}{(b^2 + \gamma_1^2 + \rho_1^2 + \rho_2^2)^{(1+f)/2}}.$$

(33) or (35) is our primary answer to the inference problem.

For the particular decision problem of selecting "interesting" factors, we need to evaluate $E(\beta_r^2)$. Since the first $n - 1$ rows of P can be described as an orthogonal transformation of the rows of W, after the latter rows have been made into unit vectors by dividing every component by $\sqrt{n}$, we see that the sum of squares of the first $n - 1$ entries in any column of P is equal to the sum of squares of entries in the corresponding column of W, divided by n. Our calculations will now be simplified if the sum of squares of entries in each column of

W is the same, equal therefore to $n(n-1)/f$. Then the sum of squares of the first $n-1$ entries in any column of P is equal to $(n-1)/f$, and that of the remaining entries is $1 - (n-1)/f = (f-n+1)/f$. Since clearly

$$(36) \quad \begin{cases} E(\gamma_r \gamma_s) = 0 \text{ for } r \neq s, \\ E(\gamma_2^2) = E(\gamma_3^2) = \ldots = E(\gamma_{n-1}^2) = E(\rho_1^2)/(n-2), \\ E(\gamma_n^2) = E(\gamma_{n+1}^2) = \ldots = E(\gamma_f^2) = E(\rho_2^2)/(f-n+1), \end{cases}$$

we obtain finally

$$(37) \quad E(\beta_r^2) = (\beta_r^2/g)^2 E(\gamma_1^2) + [(n-1)/f - (\beta_r^2/g)^2]E(\rho_1^2)/(n-2) + E(\rho_2^2)/f.$$

This can be calculated (for every value of r) as soon as $E(\gamma_1^2)$, $E(\rho_1^2)$ and $E(\rho_2^2)$ have been evaluated, and these may be found by numerical triple integration over the region $-\infty < \gamma_1 < \infty$, $0 < \rho_1 < \infty$, $0 < \rho_2 < \infty$, of four expressions in turn, namely (35) as it stands and also (35) multiplied by γ_1^2 or by ρ_1^2 or by ρ_2^2.

Our assumptions concerning X, to support the above analysis, are expressed in the following

DESIGN CONDITION ($f < n$). *The rows of X are orthogonal, the sum of squares of entries in each row is equal to n, and the sum of squares of entries in each column other than the first is equal to $n(n-1)/f$.*

The effect of the above condition can be expressed geometrically as follows. With only $n-1$ degrees of freedom available for estimating the vector (β_r) in a f-dimensional linear space, unique estimation by least squares can be carried out only in a $(n-1)$-dimensional subspace V. The condition implies that within V estimation is isotropic, i.e., has a spherical distribution of errors, and that V itself is equally inclined to each of the f co-ordinate axes. Thus the same amount of information is provided about each response coefficient.

<div align="center">REMARKS</div>

Construction of supersaturated designs

The suggested design condition for a supersaturated experiment can be satisfied for any given f and n, such that $f > n > 3$. A suitable W matrix may be constructed as follows. If $n-1$ is even, $(n-1)/2$ row pairs are taken, the mth pair consisting of a row whose rth member is $\sqrt{[(2n/f)\sin(2rm\pi/f)]}$ and a row whose rth member is $\sqrt{[(2n/f)\cos(2rm\pi/f)]}$. If $n-1$ is odd, an extra single row is included all of whose members are $\sqrt{(n/f)}$.

Hunter [5] has pointed out that the most likely occasion for a supersaturated experiment is the early part of a sequential programme in which, with f held constant, n is increased by steps, from a starting value less than $f+1$, up to $f+1$ and higher values (unless the programme is abandoned first). It can be shown that the above construction leads to a sequence of supersaturated designs, all with the same f, and with n increasing by steps of 2, such that each successive X matrix can be derived from the preceding one by adding two further rows and then rescaling the entries in accordance with our convention, so that they

satisfy (3) and have the desired sum of squares. The corresponding operation on the W matrix consists of increasing n by 2 everywhere and adding a further row pair of the kind specified above. The transformation matrix T must be appropriately modified.

Speculations

It is very likely that (30) and (37) could be well approximated by expressions far easier to calculate, and that the distributions (26) and (33) could sometimes be adequately replaced by normal distributions.

The robustness of the above analysis to non-normality of the population of β's has not been investigated, but might be expected to be fairly good.

Anyone interested in the method of "confidence intervals" could no doubt give an analogous treatment of supersaturated experiments, based on the randomization and the design condition suggested above.

SOMMAIRE

Si l'on propose une expression mathématique pour la répartition des probabilités initiales relatives à quelques paramètres, il faut faire attention aux conséquences non voulues. Quand il y a beaucoup de paramètres, il peut arriver qu'une répartition du genre nommé « conjuguée » par Raiffa et Schlaifer soit peu satisfaisante, parce qu'elle implique certaines propriétés des observations en vertu de la loi des grands nombres, lesquelles ne se vérifient pas en effet.

Comme exemple d'un problème à beaucoup de paramètres, je considère l'analyse statistique des observations faites dans une expérience factorielle, d'après une formulation due à MM. E. M. L. Beale et C. L. Mallows. Je considère en particulier l'analyse d'une expérience « sursaturée », dans laquelle le nombre des paramètres à estimer dépasse le nombre des observations. Je montre que l'analyse devient la plus facile quand le plan de l'expérience satisfait à une certaine condition. Cette condition peut s'effectuer pour un nombre quelconque de facteurs et un nombre quelconque d'observations.

REFERENCES

1. E. M. L. Beale and C. L. Mallows, On the analysis of screening experiments (Statistical Techniques Research Group, Princeton University, Technical Report No. 20, 1958). A revised version may be obtained by writing to Dr. Mallows at Bell Telephone Laboratories, Murray Hill, N.J.
2. K. V. H. Booth and D. R. Cox, Some systematic supersaturated designs, *Technometrics, 4* (1962), 489–95.
3. D. B. Duncan, Bayes rules for a common multiple comparisons problem and related Student-*t* problems, *Ann. Math. Statist., 32* (1961), 1013–33. (See especially section 5.3).
4. C. W. Dunnett, On selecting the largest of k normal population means, *J. R. Statist. Soc.,* B, *22* (1960), 1–40.
5. J. S. Hunter, Sequential factorial designs (in preparation).
6. H. Raiffa and R. Schlaifer, *Applied Statistical Decision Theory* (Boston: Harvard University Graduate School of Business Administration, 1961).
7. F. E. Satterthwaite. Random balance experimentation, *Technometrics, 1* (1959), 111–37.
8. C. M. Stein, An example of wide discrepancy between fiducial and confidence intervals, *Ann. Math. Statist., 30* (1959), 877–80.
9. ―――― Multiple regression, *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling* (Stanford University Press, 1960), 424–43.
10. G. C. Tiao and A. Zellner, Bayes' theorem and the use of prior knowledge in regression analysis (Technical report, Department of Statistics, University of Wisconsin, 1963).